

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : □@11-175081

(43)Date of publication of application : □@02.07.1999

-----  
(51)Int.Cl. □@□@G10L□@3/00□@□@  
□@ G06F□@3/00□@□@  
□@ G06F□@3/16□@□@  
□@ G06T□@1/00□@□@  
□@ G06T□@7/00□@□@

-----  
(21)Application number : □@09-341376 (71)Applicant : □@TOSHIBA CORP

(22)Date of filing : □@11.12.1997 (72)Inventor : □@SUZUKI KAORU

-----  
(54) DEVICE AND METHOD FOR SPEAKING

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a speaking device providing a user interface which inform a user of a spoken object or makes the user feel an affinity or adequacy and then improve a feeling of use.

SOLUTION: Attribute information on one or more users is estimated 1, contents to be spoken are determined 2 according to the attribute information, and speaking parameters for adjusting at least one of wording and the tone of a voice are determined 3; when the determined contents are spoken, the voice is spoken while one of the wording and the tone of the voice is changed according to the determined speaking parameters. For example, the contents to be spoken are determined according to the attribute information. The attribute information is, for example, the sex, age, and mental distance of the user. The speaking parameters are, for example, the degree of frankness of wording, the gentleness of the tone of the voice, and the speed of speaking.

-----  
LEGAL STATUS [Date of request for examination] 06.09.2000

[Date of sending the examiner's decision of rejection] 13.05.2003

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]  
□@

[Date of final disposal for application] □@

[Patent number] □@

[Date of registration] □@

[Number of appeal against examiner's decision of rejection] 2003-10631

[Date of requesting appeal against examiner's decision of rejection] 11.06.2003

[Date of extinction of right] □@

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-175081

(43) 公開日 平成11年(1999) 7月2日

(51) Int.Cl. <sup>6</sup>	識別記号	F I	
G 1 0 L 3/00		G 1 0 L 3/00	H
			R
G 0 6 F 3/00	6 0 1	G 0 6 F 3/00	6 0 1
3/16	3 3 0	3/16	3 3 0 H
G 0 6 T 1/00		15/62	3 8 0

審査請求 未請求 請求項の数 7 O L (全 14 頁) 最終頁に続く

(21) 出願番号 特願平9-341376

(22) 出願日 平成9年(1997)12月11日

(71) 出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72) 発明者 鈴木 薫

兵庫県神戸市東灘区本山南町8丁目6番26号 株式会社東芝関西研究所内

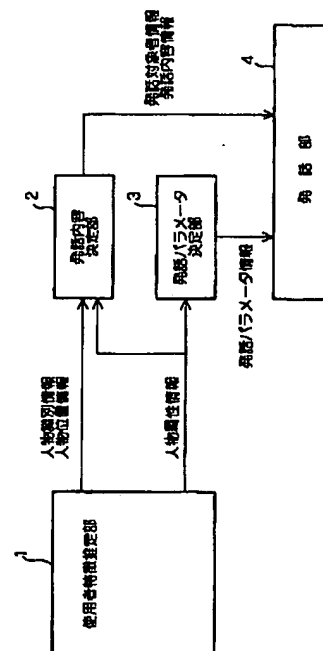
(74) 代理人 弁理士 鈴江 武彦 (外6名)

(54) 【発明の名称】 発話装置及び発話方法

(57) 【要約】

【課題】 使用者に発話対象を知らしめあるいは親近感を抱かしめあるいは適切性を感じせしめ、もって使用感を向上せしめたユーザインタフェースを提供することを可能とした発話装置を提供すること。

【解決手段】 1人以上の使用者の属性情報を推定し、前記属性情報に基づいて、発話する内容を決定し、前記属性情報に基づいて、言葉遣いおよび声の調子の少なくとも一方を調節するための発話パラメータを決定し、前記決定された内容を発話するに際して、前記決定された発話パラメータに従って言葉遣いおよび声の調子の少なくとも一方を変更しつつ発話することを特徴とする。また、例えば、前記属性情報に基づいて発話する内容を決定する。前記属性情報は、例えば、使用者の性別、年齢、心理的距離である。発話パラメータは、例えば、言葉遣いのフランク度、声の調子の優しさおよび発話の速度である。



## 【特許請求の範囲】

【請求項1】1以上の使用者の属性情報を推定する使用者特徴推定手段と、

前記属性情報に基づいて、言葉遣いおよび声の調子の少なくとも一方を調節するための発話パラメータを決定する発話パラメータ決定手段と、

所定の内容を発話するに際して、前記決定された発話パラメータに従って言葉遣いおよび声の調子の少なくとも一方を変更しつつ発声する発話手段とを具備したことを特徴とする発話装置。

【請求項2】1以上の使用者の属性情報を推定する使用者特徴推定手段と、

前記属性情報に基づいて、発話する内容を決定する発話内容決定手段と、

前記属性情報に基づいて、言葉遣いおよび声の調子の少なくとも一方を調節するための発話パラメータを決定する発話パラメータ決定手段と、

前記決定された内容が発話するに際して、前記決定された発話パラメータに従って言葉遣いおよび声の調子の少なくとも一方を変更しつつ発声する発話手段とを具備したことを特徴とする発話装置。

【請求項3】前記属性情報は、使用者の性別、年齢および心理的距離のうちの少なくとも1つの情報を含むものであることを特徴とする請求項1または2に記載の発話装置。

【請求項4】前記発話パラメータは、言葉遣いのフランク度、声の調子の優しさおよび発声の速度のうちの少なくとも1つの量からなるものであることを特徴とする請求項1ないし3のいずれか1項に記載の発話装置。

【請求項5】言葉遣いおよび声の調子の少なくとも一方の傾向を任意に変更するために、前記発話パラメータ決定手段が前記属性情報から前記発話パラメータを決定する際に用いる変更可能な内部規則を保持することを特徴とする請求項1ないし4のいずれか1項に記載の発話装置。

【請求項6】1人以上の使用者の属性情報を推定し、前記属性情報に基づいて、言葉遣いおよび声の調子の少なくとも一方を調節するための発話パラメータを決定し、

所定の内容を発話するに際して、前記決定された発話パラメータに従って言葉遣いおよび声の調子の少なくとも一方を変更しつつ発声することを特徴とする発話方法。

【請求項7】1人以上の使用者の属性情報を推定し、前記属性情報に基づいて、発話する内容を決定し、前記属性情報に基づいて、言葉遣いおよび声の調子の少なくとも一方を調節するための発話パラメータを決定し、

前記決定された内容が発話するに際して、前記決定された発話パラメータに従って言葉遣いおよび声の調子の少なくとも一方を変更しつつ発声することを特徴とする発

話方法。

## 【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、音声为主体として人間と対話するユーザインタフェースに関し、特に自然言語で人間と会話する機械の言葉遣いや声の調子を変化させるための発話装置及び発話方法に関する。

【0002】

【従来の技術】一般に、機械に命令したり機械から応答を受け取ったりといった機械と人間とのインタラクションを扱う仕掛けをユーザインタフェースと呼ぶ。ユーザインタフェースは特にコンピュータの分野で重要であり、従来からキーボードによるコマンドラインや、マウスによるグラフィカルユーザインタフェース（GUI）が実用に供されている。

【0003】GUIの次に来るインタフェースとしては、人間の話し言葉で命令できる音声インタフェースが研究され、種々提案されている。音声インタフェースでは機械が人間の話す言葉を聞き分けるための音声認識機能と、機械が人間に対して話しかけるための音声合成機能が不可欠である。これに加えて所定の目的に向けて会話の流れを制御したり、発話のタイミングなどを制御するための対話制御機能も必要である。自然言語による対話では言い回しや単語の種類、対話の進行に関して非常に大きな幅があるため、これらを正確に認識したる使い分けたりすることは難しく、話題を限定して幅を抑えることで現在実用化されつつある。

【0004】一方、人間の会話には音声の他にも表情や身振り／手振りなどの視覚的表現が存在し、かつこれが会話において重要な働きを狙っていることが知られている。電話よりも直接会って話した方がしっくりくると人が感じるのもこれに起因する。このような音声や表情などの異なる複数の表現（モダリティ）を統合的に扱えるようにするのがマルチモーダルインタフェースである。このマルチモーダルインタフェースは音声や表情／ジェスチャなどを総合的に認識／解釈し、個々の表現に曖昧性があっても全体で正確な解釈が行えるという特徴もっている。

【0005】さらに、これに人間のような顔や声を与え、使用者のパートナーとしてあたかも秘書やヘルパーのように働くものを擬人化インタフェースと呼ぶ。このインタフェースは見掛けが人間のようなだけでなく、使用者の命令を解釈して必要な手順を自分で組み立てながら処理してくれる賢いコンピュータの姿そのものである。この機械の賢さ（自律性など）を扱う技術をエージェント技術と呼ぶ。

【0006】

【発明が解決しようとする課題】以上で述べた音声インタフェース／マルチモーダルインタフェース／擬人化インタフェースは全て音声为主体にして会話を行うインタ

フェースである。しかし、これらは現在研究中のものであり、基本的な認識／合成機能の性能向上や対話モデルの検討、使用者の認証や表情／ジェスチャの認識の性能向上が中心課題となっており、実際の人間どうしの会話に迫る品質はなかなか実現できていないのが現状である。これは個々の要素を取り扱うところが難しいにもかかわらず、人間の会話というものが実に様々な要素を含んでおり、技術的にまだ扱えていない要素が多く残っているからである。

【0007】特に使用者を個人として扱うインタフェースに関しては、特定の使用者を認証して使用者専用のインタフェースエージェントを登場させるシステムと、入れ替わり訪れる複数の使用者を個々に認証しつつも単一のインタフェースエージェントに相手をさせるシステムが提案されている。前者は個人に応じた言葉遣いや声の調子を各エージェント毎に変えることは可能であるが、同一のエージェントに関しては相手に応じて言葉遣いや声の調子を適宜変えることはできない。これは後者でも同様である。これらのエージェントが使用者に応じて変えることができるのは、例えば、知っている人には「おはようございます、〇〇さん」とか、知らない人に「はじめまして」というような喋る内容の方である。すなわち、従来のエージェントは、喋る内容が同じならば大人に対しても子供に対しても同じ言葉遣いや声の調子で喋ることしかできない。この結果、従来のエージェントは複数の使用者を同時に相手にするときに、名前を呼んでから喋ったり、そちらを向いて喋ったりすることで誰に話しかけようとしているのかを使用者に知らせることができるが、使用者に応じた言葉遣いや声の調子を使い分けることで今話しかけている相手が誰であるのかを使用者に知らせることはできないのである。

【0008】また、言葉遣いや声の調子を変えることに加えて、喋る内容も使用者毎に変更することができれば、使用者はより親身に該内容を聞いてくれるものと思われるが、このような相乗効果も従来のインタフェースエージェントには期待できなかった。

【0009】本発明は、上記事情を考慮してなされたもので、同時に複数の使用者と相対した場合にも、機械が誰に話しかけているのかを各使用者に知らせ、あるいは、機械に対する親近感を該使用者に抱かしめ、あるいは、使用者に適切性を感じせしめ、もって使用感を向上せしめたユーザインタフェースを実現可能な発話装置及び発話方法を提供することを目的とする。

【0010】また、本発明は、ユーザインタフェース毎に異なる性格付けを容易に行わしめることの可能な発話装置及び発話方法を提供することを目的とする。さらに、本発明は、使用者によって発話する内容をも変えることを可能とした発話装置及び発話方法を提供することを目的とする。

【0011】

【課題を解決するための手段】本発明（請求項1）に係る発話装置は、1以上の使用者の属性情報を推定する使用者特徴推定手段と、前記属性情報に基づいて、言葉遣いおよび声の調子の少なくとも一方を調節するための発話パラメータを決定する発話パラメータ決定手段と、所定の内容を発話するに際して、前記決定された発話パラメータに従って言葉遣いおよび声の調子の少なくとも一方を変更しつつ発声する発話手段とを具備したことを特徴とする。

【0012】本発明（請求項2）に係る発話装置は、1以上の使用者の属性情報を推定する使用者特徴推定手段と、前記属性情報に基づいて、発話する内容を決定する発話内容決定手段と、前記属性情報に基づいて、言葉遣いおよび声の調子の少なくとも一方を調節するための発話パラメータを決定する発話パラメータ決定手段と、前記決定された内容が発話するに際して、前記決定された発話パラメータに従って言葉遣いおよび声の調子の少なくとも一方を変更しつつ発声する発話手段とを具備したことを特徴とする。

【0013】好ましくは、前記属性情報は、使用者の性別、年齢および心理的距離のうちの少なくとも1つの情報を含むものであってもよい。好ましくは、前記発話パラメータは、言葉遣いのフランク度、声の調子の優しさおよび発声の速度のうちの少なくとも1つの量からなるものであってもよい。

【0014】好ましくは、言葉遣いおよび声の調子の少なくとも一方の傾向を任意に変更するために、前記発話パラメータ決定手段が前記属性情報から前記発話パラメータを決定する際に用いる変更可能な内部規則を保持するようにしてもよい。

【0015】本発明（請求項6）に係る発話方法は、1人以上の使用者の属性情報を推定し、前記属性情報に基づいて、言葉遣いおよび声の調子の少なくとも一方を調節するための発話パラメータを決定し、所定の内容を発話するに際して、前記決定された発話パラメータに従って言葉遣いおよび声の調子の少なくとも一方を変更しつつ発声することを特徴とする。

【0016】本発明（請求項7）に係る発話方法は、1人以上の使用者の属性情報を推定し、前記属性情報に基づいて、発話する内容を決定し、前記属性情報に基づいて、言葉遣いおよび声の調子の少なくとも一方を調節するための発話パラメータを決定し、前記決定された内容が発話するに際して、前記決定された発話パラメータに従って言葉遣いおよび声の調子の少なくとも一方を変更しつつ発声することを特徴とする。

【0017】本発明によれば、使用者との自然言語による対話に際して、使用者の年齢／性別／心理的距離などを推定し、この推定結果に従って機械が言葉遣いや声の調子を変えることで、装置が1人の使用者とのみ相対した場合は勿論のこと、同時に複数の使用者と相対した場

合にも、各使用者は装置の言葉遣いや声の調子の差により機械が誰に話しかけているのかを各使用者に知らせ、あるいは、特定の使用者に向けた言葉遣いや声の調子により機械に対する親近感を該使用者に抱かしめ、あるいは、特定の使用者の心理的距離に適った言葉遣いや声の調子により該使用者に適切性を感じせしめることができる。

【0018】また、使用者の年齢／性別／心理的距離などを推定し、この推定結果に従って機械が発話する内容をも変えることも可能となる。よって、各使用者は装置の言葉遣いや声の調子の差により機械が誰に話しかけているのかを知ることができ、自分だけに向けた言葉遣いや声の調子で自分向けの内容と話しかけられることにより機械に対して親近感を抱き、あるいは自己との心理的距離に適った言葉遣いや声の調子に適切性を感じ、もって良好な使用感を得ることができる。

【0019】また、本発明では、上記の言葉遣いや声の調子の変わり方の傾向を様々に変えることによって、ユーザインタフェース毎に異なる性格付けを容易に行わしめることを可能とする。

【0020】よって、装置運営者にとっては、上記の言葉遣いや声の調子の変わり方の傾向を様々に変更することにより、装置毎の性格付けを統一的に行うことが可能になる。

【0021】なお、以上の各装置に係る発明は方法に係る発明としても成立し、方法に係る発明は装置に係る発明としても成立する。また、上記の発明は、相当する手順あるいは手段をコンピュータに実行させるためのプログラムを記録した機械読取り可能な媒体としても成立する。

#### 【0022】

【発明の実施の形態】以下、図面を参照しながら発明の実施の形態を説明する。以下では、本発明を適用した発話装置もしくは発話機能として、例えば音声を主体とした人間との対話あるいは人間を対象とした能動的な発話を行うようなユーザインタフェースやマルチモーダルインタフェースなどを備える装置、例えば情報提供装置、エージェント装置、能動的な広告装置等を実装する（もしくは本発明を適用した発話装置を情報提供装置、エージェント装置、能動的な広告装置等とする）ことを想定したものを例にとって本実施形態を説明する。

【0023】本発話装置もしくは発話機能はソフトウェアを用いて実現可能であり、本発話装置もしくは発話機能を実現したプログラムを計算機あるいは各種装置に実装することで、本発明に係る発話機能を搭載した各種装置あるいはサービスを実現することができる。

【0024】まず、本発明の一実施形態に係る発話装置の基本構成について説明する。図1に、本実施形態に係る発話装置の基本構成を示す。本発話装置は、使用者特徴推定部1、発話内容決定部2、発話パラメータ決定部

3、発話部4を備えている。

【0025】使用者特徴推定部1は、使用者を個々に検出して、各使用者毎に性別／年齢／心理的距離などの「人物属性情報」を推定するとともに、必要なら各使用者の位置を表す「人物位置情報」や各使用者の名前などを表す「人物識別情報（未知人物を示す識別情報を含む）」を判別する。

【0026】発話内容決定部2は、使用者特徴推定部1により判別された各使用者の人物識別情報と人物属性情報と人物位置情報の全てあるいはその一部をもとにして、どの使用者に何を話すかを表す「発話対象者情報」と「発話内容情報」を生成する。

【0027】発話パラメータ決定部3は、使用者特徴推定部1により推定された発話対象者の人物属性情報に応じた「言葉遣い」および／または「声の調子」を制御する発話パラメータ情報を算出する。

【0028】発話部4は、発話内容決定部2で決定された発話対象者情報に従って必要ならエージェント映像の顔向きや視線（装置に付随する人形やこれに相当するものなどあるいは装置本体や表示画面などに関する可動部分が電子的に制御可能な場合には人形等の顔向きや視線あるいは装置本体や表示画面等の対面方向など）を制御するとともに、発話内容決定部2で決定された内容を発話パラメータ決定部3で決定された発話パラメータ情報に従って発話出力する。

【0029】次に、本実施形態に係る発話装置における基本的な処理について説明する。図2に、本実施形態に係る発話装置における処理の流れの一例を示す。ステップS1の使用者特徴推定処理では、各使用者毎に性別／年齢／心理的距離などの人物属性情報を推定するとともに、必要なら各使用者の位置を表す人物位置情報や各使用者の名前などを表す人物識別情報（未知人物を示す識別情報を含む）を判別する。

【0030】ステップS2の発話内容決定処理では、ステップS1にて判別された各使用者の人物識別情報と人物属性情報と人物位置情報の全てあるいはその一部をもとにして、どの使用者に何を話すかを表す発話対象者情報と発話内容情報を生成する。

【0031】ステップS3の発話パラメータ決定処理では、ステップS1にて推定された発話対象者の人物属性情報に応じた言葉遣いおよび／または声の調子を制御する発話パラメータ情報を算出する。

【0032】ステップS4の発話処理では、ステップS2にて決定された発話対象者情報に従って必要ならエージェント映像の顔向きや視線（装置に付随する人形やこれに相当するものなどあるいは装置本体や表示画面などに関する可動部分が電子的に制御可能な場合には人形等の顔向きや視線あるいは装置本体や表示画面等の対面方向など）を制御するとともに、ステップS2にて決定された内容をステップS3にて決定された発話パラメータ

情報に従って発話出力する。

【0033】以下では、使用者特徴推定部1、発話内容決定部2、発話パラメータ決定部3、発話部4の順に、より詳しい構成例を用いて説明していく。まず、使用者特徴推定部1についてより詳しく説明する。

【0034】図3に、使用者特徴推定部1の構成例を示す。図3に示されるように、この使用者特徴推定部1は、人物検出辞書11、人物検出部12、人物照合辞書13、人物照合部14、心理的距離推定部15、性別照合辞書16、性別推定部17、年齢照合辞書18、年齢推定部19を備えている。

【0035】人物検出辞書11は、人物検出用テンプレートを格納する。人物検出部12は、装置前面の使用者と相対する空間の画像を入力し、該画像中から上記の人物検出用テンプレートと十分に類似する部分を検出して人物位置情報として出力するとともに、検出された人物の画像部分を人物領域画像として出力する。

【0036】人物照合辞書13は、個人を照合するための既知人物毎の人物照合用テンプレートを格納する。人物照合部14は、人物検出部12により得られた人物領域画像を上記の人物照合用テンプレートと照合して最高かつ十分に類似するテンプレートを選択し、該テンプレートに対応付けられた人物識別情報を出力する。なお、人物照合部14は、人物領域画像を人物照合用テンプレートと照合した結果、十分に類似するテンプレートを発見できなかった場合には未知人物であることを示す人物識別情報を生成する。

【0037】心理的距離推定部15は、上記の人物識別情報に基づいて、既知人物毎の当該装置の装置使用回数を更新するとともに、該更新された装置使用回数に応じた心理的距離を算出して出力する。この心理的距離は、この場合どれぐらい装置に親しんだかを示す値であり、装置使用回数に対して単調減少する量、例えば、装置使用回数に反比例する量である。

【0038】性別照合辞書16は、性別照合用テンプレートを格納する。性別推定部17は、人物検出部12により得られた人物領域画像を上記の性別照合用テンプレートと照合して最高に類似するテンプレートを選択し、該テンプレートに対応付けられた性別を出力する。

【0039】年齢照合辞書18は、年齢照合用テンプレートを格納する。年齢推定部19は、人物検出部12により得られた人物領域画像を上記の年齢照合用テンプレートと照合して最高に類似するテンプレートを選択し、該テンプレートに対応付けられた年齢を出力する。

【0040】なお、性別推定部17は、人物照合部14と同様に、最高かつ十分に類似するテンプレートが発見できた場合のみ該テンプレートに対応付けられた性別を出力し、十分に類似するテンプレートが発見できなかった場合は、性別不定であることを示す情報を出力するようにしてもよい。同様に、年齢推定部19は、最高かつ

十分に類似するテンプレートが発見できた場合のみ該テンプレートに対応付けられた年齢を出力し、十分に類似するテンプレートが発見できなかった場合は、年齢不定であることを示す情報を出力するようにしてもよい。

【0041】なお、心理的距離推定部15の出力する心理的距離と性別推定部17の出力する性別と年齢推定部19の出力する年齢とがすなわち人物属性情報である。また、人物検出辞書11／人物照合辞書13／性別照合辞書16／年齢照合辞書18に格納されるテンプレートは各々対応するカテゴリ（人間／個人別／男女／年齢別）毎に十分数取得された元画像から統計的に求められる各カテゴリの特徴を表現した情報である。

【0042】次に、発話内容決定部2についてより詳しく説明する。図4に発話内容決定部2の構成例を示す。図4に示されるように、この発話内容決定部2は、制御規則データベース21、発話制御部22、選択規則データベース23、話題データベース24、発話内容選択部25から構成される。

【0043】制御規則データベース（制御規則DB）21は、発話の状態を制御するための制御規則を格納する。発話制御部22は、上記の制御規則に従って発話の状態を制御する内容選択指令を生成する。内容選択指令は、例えば、「〇〇さんに挨拶をせよ」とか「本題を述べよ」などの発話すべき内容の比較的大まかな指示を表す。

【0044】選択規則データベース（選択規則DB）23は、発話すべき内容を決定するための選択規則を格納する。話題データベース（話題DB）24は、発話すべき話題を格納する。

【0045】発話内容選択部25は、話題DB24から、発話制御部22の出力する内容選択指令に合致する内容を、選択規則DB23に格納される選択規則に基づいて読み出すとともに、どの使用者をターゲットとして話すのかを示す発話対象者情報を生成する。

【0046】ここで、制御規則DB21に格納する制御規則について説明する。制御規則は大別すると発話構成規則と発話対象規則の2種類からなる。前者は発話の進行を司る規則であり、後者は発話対象者の決定を司る規則である。

【0047】最初に、発話構成規則について詳しく説明する。発話構成規則は、当該装置が最初に使用者を検出してから所定の話者を喋り終わるまでの発話の状態を遷移させる規則であり、挨拶や本題の発話順序を定めるものである。

【0048】図5に、発話構成規則の一例を示す。図5に例示する発話構成規則では、待ち受け状態から使用者を検出するとまず開始挨拶（「こんにちわ」など）が発話され、次に該使用者（あるいは使用者の集団）に対して所定の本題（何かの売り込みや自己紹介など）が順次発話される。本題の発話は喋るべき本題がなくなるま

で繰り返行われる。全ての本題を話し終えると、終了挨拶（「どうもありがとうございました」など）が発話されて発話が終了する。もし、途中で使用者がいなくなれば、その時点で本題などの発話は中止され、中断挨拶（「お忙しいのですか？」など）が発話されて待ち受け状態に移る。

【0049】続いて、発話対象規則について詳しく説明する。発話対象規則は、発話構成規則が指示する発話の最中にどのように話しかける相手（発話対象者）を変えていくかを決定する規則であり、文節単位あるいは文単位で、対象者の位置と属性情報に応じて、エージェント映像等の視線／顔向きや、発話の際の言葉遣い／声の調子などを変えるために用いられる。

【0050】発話対象規則には、発話開始時に適用される開始時巡回規則と、その後続く発話中盤に適用される中盤巡回規則の2つを用意している。例えば、次に例示する開始時巡回規則のように、最も近くに位置する人物から順に挨拶してゆき、最後はランダムに巡回する。もし、子供がいる場合には優先的に子供から発話対象を巡回させる。

【0051】図6に、発話対象規則のうちの開始巡回規則の一例を示す。図6に例示した開始時巡回規則では、親子2人連れの利用者がいる場合、まず子供に対して「こんにちは、ボク。」と挨拶し、次いで親に対して「私は、××です。」と自己紹介する。なお、この例では巡回を文単位としている。挨拶文の文数は自由であるが、挨拶文を言い終えてもさらに未巡回の利用者がいる場合には、開始時巡回規則の適用を終了して本題に入る。

【0052】さて、本題に入ると中盤巡回規則が適用される。図7に、発話対象規則のうち中盤巡回規則の一例を示す。図7に例示した中盤巡回規則では、話題に適合する人物および話題に熱心な人物に対して優先的に話しかける。話題には各々最も適した人物のカテゴリ（性別、年齢）があるので、ある人物の属性情報のうちいくつかが話題に適合するかで該人物の適合度（適合する属性数）を計算する。また、ある人物が当該装置への積極的な接近を試みたり、当該装置に対して顔をずっと向けている場合には該人物を熱心な人物であると判定して熱心度（0/1）を計算する。次に適合度と熱心度の加重和を巡回優先度とし、この巡回優先度にしたがって対象者を巡回する。1人の人物を対象とする時間は数秒から数十秒程度であり巡回優先度に比例する。

【0053】次に、選択規則DB23に格納する選択規則について説明する。選択規則は、内容選択指令に従って発話内容を選択するための規則である。図8に選択規則の一例を示す。

【0054】図8に例示した選択規則では、開始挨拶、終了挨拶、中断挨拶は大人／子供共通の内容が選択されるが、本題については大人と子供で異なる内容が選択さ

れる。この他にも、本題の選択規則としては性別や心理的距離による区別も考えられる。

【0055】次に、話題DB24に格納される情報について説明する。話題DB24の内容は標準的な言葉遣いで記述された文章である。本実施形態に係る発話装置を能動的な広告装置として適用した場合の話題DB24の内容を以下に例示する。

【0056】〔話題の例〕

対象=既知人物：種別=開始挨拶：文=「こんにちは○さん。」

対象=未知人物：種別=開始挨拶：文=「はじめまして。私は××です。」

対象=全カテゴリ：種別=終了挨拶：文=「どうもありがとうございました。」

対象=全カテゴリ：種別=中断挨拶：文=「お忙しいのですか？ では、またお会いしましょう。」

対象=大人：種別=本題：文=「さて、最近…」（海外旅行の案内）

対象=大人：種別=本題：文=「今、一番…」（新作映画の案内）

対象=子供：種別=本題：文=「今度、…」（新作ゲームの案内）

対象=大人：種別=本題：文=「今、一番…」（新作映画の案内）

対象=全カテゴリ：種別=本題：文=「新しいアトラクション…」（テーマパークの案内）

対象=大人かつ女性：種別=本題：文=「この秋の新品は…」（化粧品品の案内）

各話題は、対象者のカテゴリを規定する対象者リスト

と、挨拶や本題などの話題の種別を規定する識別情報

と、実際の文とからなっている。発話内容選択部25

は、選択規則の適用に従って定められる話題の対象者を対象者リストに含む話題を検索し、発話の完了していない話題を順次選び出すように働く。選び出された話題に記載される文が発話内容決定部2の出力する発話内容情報である。

【0057】ただし、未知人物の場合には、対象=既知人物とする文、例えば

対象=既知人物：種別=開始挨拶：文=「こんにちは○さん。」

は選択不可とする。

【0058】なお、話題DB24の文には、補足資料（カタログ写真など）を適所にリンクし、リンク箇所を発話するときには該当する補足資料が同時に提示されるようにしてもよい。また、各文にはそのときのエージェントの身体動作を規定する情報も付加し、例えば、挨拶時には会釈動作、補足資料提示時には該資料の指し示し動作などをエージェントが行えるようにしてもよい。

【0059】次に、発話パラメータ決定部3についてより詳しく説明する。図9に、発話パラメータ決定部3の

構成例を示す。図9に示されるように、この発話パラメータ決定部3は、標準特性データベース31、標準パラメータ設定部32、個別特性データベース33、個別パラメータ設定部34から構成される。

【0060】標準特性データベース（標準特性DB）31は、対象者のカテゴリに応じて標準的な発話パラメータを決定するための標準調整セットを格納する。標準パラメータ設定部32は、使用者特徴推定部1の出力する人物属性情報と上記の標準調整セットに基づいて、標準的な発話パラメータ情報を決定する。

【0061】個別特性データベース（個別特性DB）33は、装置個々の性格を表出するための個別特性を格納する。個別パラメータ設定部34は、標準パラメータ設定部32の出力する標準パラメータ情報を受け、これに

〔基本パラメータセット〕

言葉遣いのフランク度=50（標準値）

声の調子の優しさ度=50（標準値）

話す速さ=50（標準値）

〔子供向け標準調整セット〕

言葉遣いのフランク度 +50（極めてくれた感じに調整）

声の調子の優しさ度 +50（極めて優しい感じに調整）

話す速さ -30（ゆっくりした感じに調整）

〔大人向け標準調整セット〕

言葉遣いのフランク度 ±0（標準値維持）

声の調子の優しさ度 ±0（標準値維持）

話す速さ ±0（標準値維持）

〔老人向け標準調整セット〕

言葉遣いのフランク度 ±0（標準値維持）

声の調子の優しさ度 +20（やや優しい感じに調整）

話す速さ -30（ゆっくりした感じに調整）

〔心理的距離の近い人物向け標準調整セット〕

言葉遣いのフランク度 +50（極めてくれた感じに調整）

声の調子の優しさ度 ±0（標準値維持）

話す速さ ±0（標準値維持）

〔心理的距離の遠い人物向け標準調整セット〕

言葉遣いのフランク度 -50（極めてあらたまった感じに調整）

声の調子の優しさ度 ±0（標準値維持）

話す速さ ±0（標準値維持）

これらはどの装置にも共通する対象者のカテゴリ毎に決まる標準的な言葉遣いと声の調子を設定する規則である。これを次段の個別パラメータ設定部34により、装置毎の個性を与えられた言葉遣いと声の調子を生成可能

〔カテゴリに共通の個別特性〕

言葉遣いのフランク度 +10（誰に対してもくれた感じで喋る傾向）

声の調子の優しさ度 +10（誰に対しても優しく喋る傾向）

話す速さ -20（誰に対してもゆっくり話す傾向）

〔カテゴリに固有の個別特性〕

if（対象者が子供） then 声の調子の優しさ +10（子供には特に優しい傾向）

カテゴリに共通の個別特性とは対象者のカテゴリに依存

個別DB33に格納される個別特性による偏向を加えた発話パラメータ情報を生成する。

【0062】標準特性DB31には1つの基本パラメータセットがあり、これに対する調整量を規定した対象者のカテゴリ毎の標準調整セットが格納される。具体的には、子供向けの標準調整セットや女性向けの標準調整セットなどが格納されている。標準パラメータ設定部32は与えられた対象者のカテゴリから該当する標準調整セットを読み出し、これに基づいて基本パラメータセットを調整してカテゴリに応じた標準パラメータを生成する。

【0063】基本パラメータセットと標準調整セットの具体例を以下に示す。

なパラメータに変換する。個別特性DB33にはこの個性を付けるための個別特性が格納されている。

【0064】個別特性の具体例を以下に示す。

しない装置の個性であり、誰に対しても同じ傾向で喋る



ことを意味する。カテゴリに固有の個別特性は、例えば子供には優しい装置というように対象者の特定カテゴリに対して効力を発揮する。このように個別特性を設けることで、共通の枠組みで装置の個性を演出することが可能になる。

【0065】次に、発話部4についてより詳しく説明する。図10に、発話部4の構成例を示す。図10に示されるように、この発話部4は、文変換規則データベース41、発話文変換部42、聴覚情報生成部43、視覚情報生成部44から構成される。

【0066】文変換規則データベース（文変換規則DB）41は、発話内容決定部2の出力する発話内容情報（標準的な言葉遣いで記述された読み上げ文）を所定の言葉遣いの文に変換するための規則を格納する。

【0067】発話文変換部42は、上記の発話内容情報を発話内容決定部2の出力する発話パラメータに従って上記の文変換規則DB41を参照しつつ所定の言葉遣いの文に変換する。

【0068】聴覚情報生成部43は、発話文変換部42の出力する発話文情報を上記の発話パラメータに従って実際の音声に変換出力する。視覚情報生成部44は、発話対象者情報に応じた顔向き／視線、および発話内容情報の付加情報に応じた動作を与えられたエージェント映像を生成し（あるいは電子的に制御可能な人形等を制御し）、必要ならば発話内容情報にリンクされた補足資料を同時に映像化する。

【0069】聴覚情報生成部43は、音声を生成出力するとともに、該音声に対応した同期情報を視覚情報生成部44に受け渡す。同期情報を受け渡された視覚情報生成部44は該同期情報に応じてエージェントの口の開閉を行う。

【0070】なお、このとき発話対象者情報、発話パラメータ情報、発話内容情報とそれらに基づいて変換される発話文情報の関係は例えば次のようなものである。

〔発話対象者情報の例〕

氏名＝鈴木薫

性別＝男性

年齢＝8才

心理的距離＝60

位置＝××

〔発話パラメータ情報の例〕

言葉遣いのフランク度＝100

声の調子の優しさ＝100

話す速さ＝40

〔発話文情報（変換前）の例〕

「こんにちわ〇〇さん。新しいアトラクションが充実した…をご紹介します。」

〔発話文情報（変換後）の例〕

「こんにちわ薫くん。新しいアトラクションがいったいの…を紹介するね。」

これを行うために、文変換規則DB41には以下のような変換規則が記述されている。変換規則には以下に例示するように呼称代入規則、文末変換規則、言い換え規則がある。図11に呼称代入規則の一例を示す。図12に文末変換規則の一例を示す。図13に言い換え規則の一例を示す。

【0071】上記の呼称代入規則は発話内容情報に対象者などの名前を入れなければならないときに、該箇所適切に呼称を代入するための規則である。また、文末変換規則はフランク度に応じて文末を適切に偏向するための規則である。言い換え規則は特に子供にとって難しい単語を平易な単語に置き換えるための規則である。

【0072】文末変換規則と言い換え規則はその適用に際して、原文（発話内容情報）で用いられる各単語のフランク度や難易度を検査しなければならない。この検査のためには許容値リストが用いられる。許容値リストは各単語のフランク度や難易度の適用範囲を規定する情報である。この範囲を逸脱している単語は逸脱しない別の単語に置き換えられなければならない。この置き換えのために同義語リストが用いられる。同義語リストは各単語の同義語をまとめた情報である。許容値を逸脱した原文の単語は逸脱しない別の単語をこの同義語リストから検索されて置き換えられる。置き換え可能な同義語が複数存在する場合には、ランダムに1つを決めて置き換える。このようにすることで、同じ単語が文中に繰り返し現れることを回避する。

【0073】許容値リストと同義語リストの例を以下に示す。なお、許容値リストと同義語リストは文変換規則とともに文変換規則DB41に格納される。

〔許容値リストの例〕

「するよ」「するね」「させてね」「しないよ」など：フランク度＞80

「します」「しません」など：80≧フランク度≧20

「させていただきます」「致します」「致しません」「ご（接頭語）」など：20＞フランク度

「充実」など：難易度＞50

「いっぱい」「たっぷり」など：50≧難易度

「はじめまして」など：難易度＞50

「こんにちは、ボク」など：50≧難易度

〔同義語リストの例〕

「するよ」「するね」「させてね」「します」「させていただきます」「致します」など

「しないよ」「しません」「致しません」など

音声生成部43は以上のようにして生成された発話文情報を発話パラメータの指定する声の調子と速さで音声に変換する。各々の文末語には語尾の上げ下げの情報が付属している。例えば、「するね」の語尾は上がり、「します」の語尾は下がる。これに加えて、声の調子が優しい場合には、音声パワーのダイナミックレンジを狭くしてきつく聞こえないようにし、さらに文末語の語尾を延

ばす処理を加えて優しさを表現する。

【0074】なお、本発明に係る発話装置や発話方法は以上の実施形態に限定されるものではない。例えば、発話内容を利用者のカテゴリに応じて変更せず、所定の内容を利用者に応じた言葉遣いや声の調子で発話するようにしてもよい。

【0075】また、人物検出辞書11、人物照合辞書13、性別照合辞書16、年齢照合辞書18、制御規則DB21、選択規則DB23、話題DB24、標準特性DB31、個別特性DB33、文変換規則DB41に格納される各種情報を装置外部から入れ替え可能として、外部からの操作により様々に調整可能としてもよい。

【0076】また、利用者特徴推定部1が、画像のみならず音声などの他の情報を単独あるいは同時に用いて人物識別情報、人物位置情報、人物属性情報を推定することも可能である。

【0077】また、音声でなく文字により言葉遣いを変更した出力を可能としてもよい。ところで、本発明に係る発話装置や発話方法を実現する情報（例えばプログラム）を図14に示すように記録媒体51に記録し、該記録した情報を該記録媒体51を経由して装置52に適用したり、通信回線を経由して装置53に適用することも可能である。本発明は、上述した実施の形態に限定されるものではなく、その技術的範囲において種々変形して実施することができる。

【0078】

【発明の効果】本発明によれば、装置が複数の使用者と同時に相対した場合に、各使用者は装置の言葉遣いや声の調子の差により機械が誰に話しかけているのかを知ることができ、自分だけに向けた言葉遣いや声の調子で自分向けの内容を話しかけられることにより機械に対して親近感を抱き、あるいは自己との心理的距離に適った言葉遣いや声の調子に適切性を感じ、もって良好な使用感を得ることができる。

【0079】また、装置運営者にとっては、上記の言葉遣いや声の調子の変わり方の傾向を様々に変更することにより、装置毎の性格付けを統一的に行うことが可能になる。

【図面の簡単な説明】

【図1】本発明の一実施形態に係る発話装置の基本構成を示す図

【図2】同実施形態に係る発話装置における処理手順の一例を示すフローチャート

【図3】使用者特徴推定部の構成の一例を示す図

【図4】発話内容決定部の構成の一例を示す図

【図5】発話構成規則の一例を示す図

【図6】開始巡回規則の一例を示す図

【図7】中盤巡回規則の一例を示す図

【図8】選択規則の一例を示す図

【図9】発話パラメータ決定部の構成の一例を示す図

【図10】発話部の構成の一例を示す図

【図11】呼称代入規則の一例を示す図

【図12】文末変換規則の一例を示す図

【図13】言い換え規則の一例を示す図

【図14】記録媒体等による実施形態を説明するための図

【符号の説明】

1…利用者特徴推定部

2…発話内容決定部

3…発話パラメータ決定部

4…発話部

11…人物検出辞書

12…人物検出部

13…人物照合辞書

14…人物照合部

15…心理的距離推定部

16…性別照合辞書

17…性別推定部

18…年齢照合辞書

19…年齢推定部

21…制御規則データベース

22…発話制御部

23…選択規則データベース

24…話題データベース

25…発話内容選択部

31…標準特性データベース

32…標準パラメータ設定部

33…個別特性データベース

34…個別パラメータ設定部

41…文変換規則データベース

42…発話文変換部

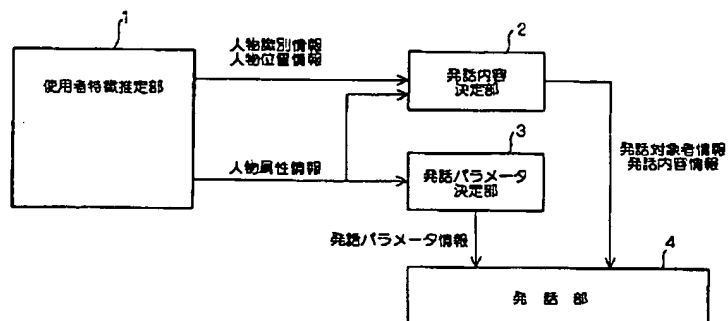
43…聴覚情報生成部

44…視覚情報生成部

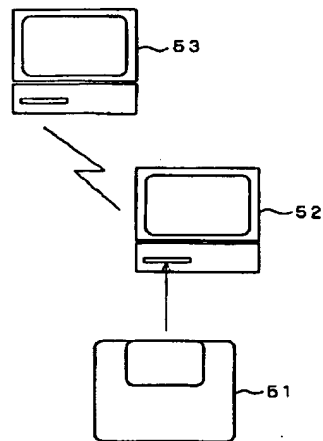
51…記録媒体

52、53…装置

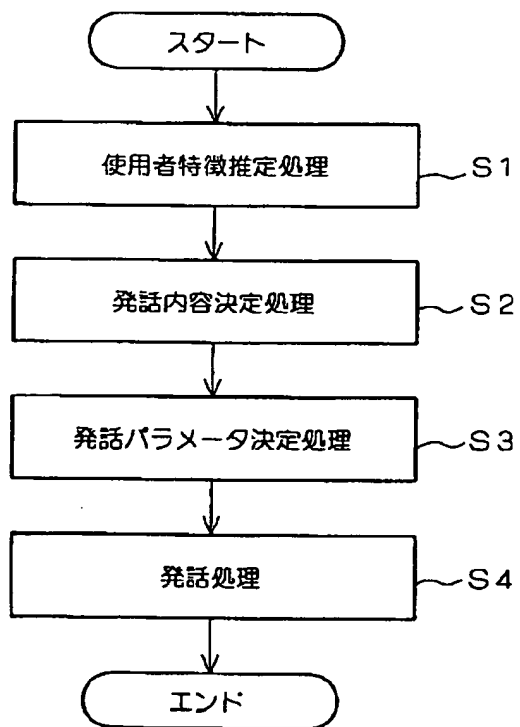
【図1】



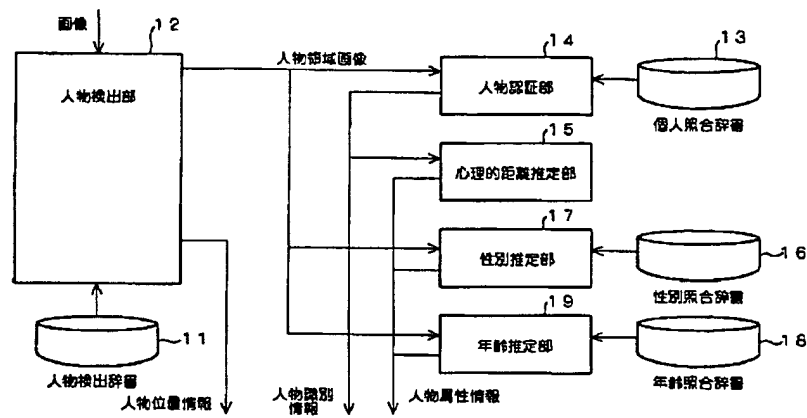
【図14】



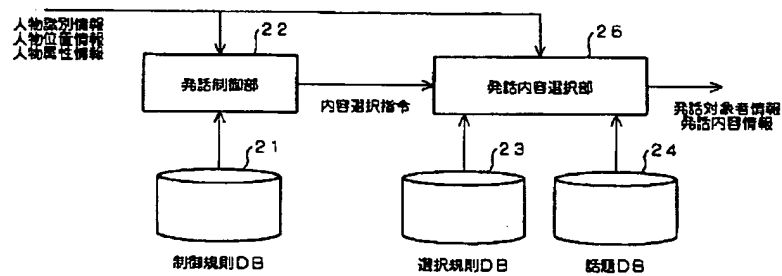
【図2】



【図3】



【図4】



【図5】

## 【制御規則の例：発話構成規則の例】

```

if (発話中でない) then {
  if (新しい使用者がいる) then 開始挨拶開始 (開始挨拶を指令)
}
if (開始挨拶終了) then 最初の本題を開始 (先頭本題を指令)
if (現在の本題終了) then {
  if (全ての本題を終了) then 終了挨拶開始 (終了挨拶を指令)
  if (別の本題が残っている) then 次の本題を開始 (次本題を指令)
}
if (終了挨拶終了) then 発話終了 (発話終了を指令)
if (発話中であるand 使用者がいない) then 中断挨拶開始 (中断挨拶を指令)
if (中断挨拶終了) then 発話終了 (発話終了を指令)

```

【図6】

```

【制御規則の例：発話対象規則：開始時巡回規則の例】
if (発話開始時) then {
  if (子供がいる) then 最も近い子供を対象者とする
  if (子供がいない) then 最も近い大人を対象者とする
}
if (現在の対象者は子供) then {
  if (子供は他にもいる) then 次に近い子供を対象者とする
  if (子供は他にいない) then 最も近い大人を対象者とする
}
if (現在の対象者は大人) then {
  if (大人は他にもいる) then 次に近い大人を対象者とする
  if (大人は他にいない) then ランダムに巡回する
}
if (挨拶終了) then 中盤巡回規則に移行する

```

【図7】

```

【制御規則の例：発話対象規則：中盤巡回規則の例】
if (巡回優先度がゼロでない人物がいる) {
  if (中盤開始時) then 最も巡回優先度の高い人物を対象者とする
  if (巡回優先度がゼロでない人物は他にもいる)
    then 次に巡回優先度の高い人物を対象者とする
  if (巡回優先度がゼロでない人物は他にいない)
    then 巡回優先度がゼロでない人物から対象者をランダムに決める
}
if (巡回優先度がゼロでない人物がいない) 対象者をランダムに決める

```

【図8】

## 【選択規則の例】

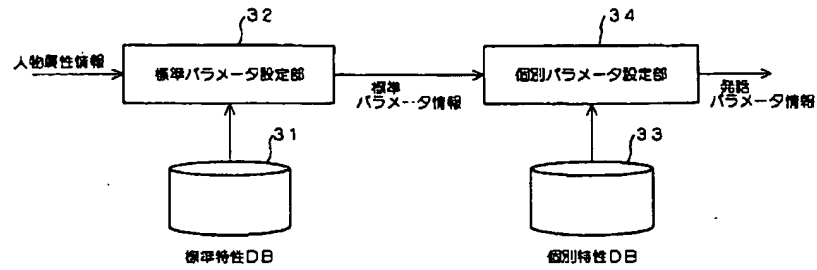
```

if (開始挨拶指令) then 開始挨拶を話題DBから選択
if (終了挨拶指令) then 終了挨拶を話題DBから選択
if (中断挨拶指令) then 中断挨拶を話題DBから選択

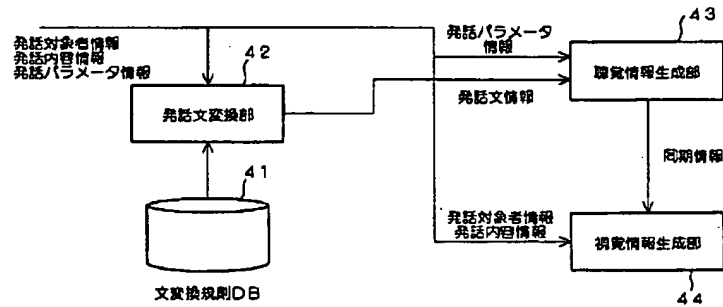
if (本題指令) then {
  if (使用者は大人だけ) then 大人を対象に含む本題を話題DBから選択
  if (使用者は子供を含む) then 子供を対象に含む本題を話題DBから選択
}

```

【図9】



【図10】



【図11】

【変換規則：呼称代入規則の例】

```

if (氏名が入る箇所) then
  if (年齢<10) then
    if (男性) then 名前+「くん」を使う
    else          名前+「ちゃん」を使う
  else
    if (フランク度>80) then 名前+「さん」を使う
    else
      if (フランク度<20) then 名字+「さま」を使う
      else                    名字+「さん」を使う
  }
}

```

【図12】

【交換規則：文末交換規則の例】

```

if (文末) then {
  if (文末語がフランク度の許容値を逸脱) then {
    フランク度の許容値にある同義語をランダムに選択して置換
  }
}

```

【図13】

【交換規則：言い替え規則の例】

```

if (子供) then {
  if (単語の難易度が子供の許容値 (≦50) を逸脱) then {
    難易度の許容値にある同義語をランダムに選択して置換
  }
}

```

---

フロントページの続き

(51)Int.Cl.<sup>6</sup>

G06T 7/00

識別記号

FI

G06F 15/62

465K

## \* NOTICES \*

JPO and NCIPI are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.

2. \*\*\*\* shows the word which can not be translated.

3. In the drawings, any words are not translated.

## [Claim(s)]

[Claim 1] A user description presumption means to presume one or more users' attribute information, and a voice synthesis parameter decision means to determine the voice synthesis parameter for adjusting at least one side of the tune of wording and voice based on said attribute information, voice synthesis equipment characterized by providing a voice synthesis means to utter facing uttering the predetermined contents and changing at least one side of the tune of wording and voice according to said determined voice synthesis parameter.

[Claim 2] A user description presumption means to presume one or more users' attribute information, and a contents decision means of voice synthesis to determine the contents which speak based on said attribute information, A voice synthesis parameter decision means to determine the voice synthesis parameter for adjusting at least one side of the tune of wording and voice based on said attribute information, voice synthesis equipment characterized by providing a voice synthesis means to utter facing uttering said determined contents and changing at least one side of the tune of wording and voice according to said determined voice synthesis parameter.

[Claim 3] Said attribute information is voice synthesis equipment according to claim 1 or 2 characterized by being a thing including at least one information in a user's sex, age, and psychic distance.

[Claim 4] Said voice synthesis parameter is voice synthesis equipment given in claim 1 characterized by being what consists of at least one amount in the tenderness of the tune of voice, and the rate of voice synthesis thru/or any 1 term of 3 whenever [ flank / of wording ].

[Claim 5] voice synthesis equipment given in claim 1 characterized by holding the internal rule which is used in case said voice synthesis parameter decision means determines said voice synthesis parameter from said attribute information, and which can be changed in order to change one [ at least ] inclination of the tune of wording and voice into arbitration thru/or any 1 term of 4.

[Claim 6] The voice synthesis approach characterized by uttering presuming one or more users' attribute information, determining the voice synthesis parameter for adjusting at least one side of the tune of wording and voice based on said attribute information, facing uttering the predetermined contents, and changing at least one side of the tune of wording and voice according to said determined voice synthesis parameter.

[Claim 7] Presume one or more users' attribute information, and the contents which speak are determined based on said attribute information. Based on said attribute information, determine the voice synthesis parameter for adjusting at least one side of the tune of wording and voice, and it faces uttering said determined contents. The voice synthesis approach characterized by uttering changing at least one side of the tune of wording and voice according to said determined voice synthesis parameter.

-----  
[Translation done.]  
-----

-----  
DETAILED DESCRIPTION  
-----

[Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention relates to the voice synthesis equipment and the voice synthesis approach for changing the wording of a machine and the tune of voice of talking with human being especially with natural language, about the user interface which converses with human being by making voice into a subject.

[0002]

[Description of the Prior Art] The mechanism in which the interaction of the machine and human being of generally being as receiving a response from a machine \*\*\*\* [ and ] is treated is called a user interface. [ ordering a machine ] Especially the user interface is important in the field of a computer, and practical use is presented with the command line by the keyboard, and the graphical user interface (GUI) by the mouse from the former.



[0003] as the interface which comes to the degree of GUI -- human being's talk -- the voice interface which can be ordered by means of language is studied, and many things are proposed. A speech recognition function for a machine to recognize the language about which human being speaks in a voice interface, and a speech synthesis function for a machine to address to human being are indispensable. In addition, the dialogue control function for controlling the flow of conversation or controlling the timing of voice synthesis etc. towards the predetermined purpose, is also required. In the dialogue by natural language, since there is very big width of face about the class of an expression or word, and advance of a dialogue, these are recognized correctly, slack usage part beam squirrel \*\*\*\*\* is difficult, and current utilization is being carried out by limiting subject and stopping width of face.

[0004] It is known that the visual expression of expression, a gesture/gesture, etc., etc. other than voice exists in human being's conversation, and this is aiming at important work in conversation on the other hand. If the direction said as the direct meeting rather than the telephone comes exactly, that people sense also originates in this. A multi-modal interface enables it to treat two or more different expressions (modality), such as such voice and expression, integrative. This multi-modal interface recognizes / interprets voice, expression/gesture, etc. synthetically, and even if ambiguity is shown in each expression, it has the description that an exact interpretation can be performed on the whole.

[0005] Furthermore, the face and voice like human being are given to this, and what works just like a secretary or a helper as a user's partner is called a personification interface. This interface is the figure of the wise computer processed while appearance interprets an instruction of a user and assembles [ not only like / appearance / human being but ] a required procedure by itself itself. The technique of treating the cleverness (autonomy etc.) of this machine is called an agent technique.

[0006]

[Problem(s) to be Solved by the Invention] All voice interface / multi-modal interface / personification interfaces that were described above are interfaces which talk by making voice into a subject. However, these are the things under current research and the present condition is that the quality for which the fundamental improvement in the engine performance of recognition/composition function and examination of a dialogue model, and the improvement in the engine performance of a user's authentication and recognition of expression/gesture press actual human beings' conversation by having been a main technical problem is not easily realizable. Although the place which deals with each element is difficult for this, a thing called human being's conversation contains very various elements, and it is because many elements which have not been carried yet technically remain.

[0007] About the interface which treats especially a user as an individual, the system in which attest a specific user and the interface agent only for users is made to appear, and the system to which a single interface agent is made to do a partner although two or more users who interchange and visit are attested separately are proposed. Although it is possible for the former to change the tune of the wording according to an individual or voice for every agent, about the same agent, the tune of wording or voice is unchangeable suitably according to a partner. As for this, the same is said of the latter. these agents can change according to a user -- for example, it is the direction of contents which say to those who know at "Mr. good morning and OO" and those who do not know, "how do you do" and which talk. That is, if the conventional agent has the the same contents which talk, he can only do talking at the tune of the same wording and voice also to a child also to an adult. Consequently, although the conventional agent can inform a user of whom it is going to speak to by talking after calling an identifier, or turning to and talking there when making two or more users into a partner at coincidence, it cannot talk now by using the tune of the wording and the voice according to a user properly, and a user cannot be told about \*\* or who a \*\*\*\*\* partner is.

[0008] moreover, the thing for which the tune of wording and voice is changed -- in addition, although it was thought that a user heard these contents kindlier when the contents which talk could also be changed for every user, such the synergistic effect was not expectable to the conventional interface agent, either.

[0009] This invention aims at offering the voice synthesis equipment and the voice synthesis approach the user interface which make each user know whom the machine is speaking to, or this user was made to have a sense of closeness over a machine, or sensed felicity, cheated out of it and had it in the user, and made a feeling of use improve also when it was made in consideration of the above-mentioned situation and two or more users were faced at coincidence is realizable.

[0010] Moreover, this invention aims at offering the possible voice synthesis equipment and the voice synthesis approach of making different characterization for every user interface perform easily. Furthermore, this invention aims at offering the voice synthesis equipment and the voice synthesis approach which made it possible to also change the contents which speak by the user.

[0011]

[Means for Solving the Problem] A user description presumption means by which the voice synthesis equipment concerning this invention (claim 1) presumes one or more users' attribute information, An voice synthesis parameter

decision means to determine the voice synthesis parameter for adjusting at least one side of the tune of wording and voice based on said attribute information, It is characterized by providing an voice synthesis means to utter facing uttering the predetermined contents and changing at least one side of the tune of wording and voice according to said determined voice synthesis parameter.

[0012] A user description presumption means by which the voice synthesis equipment concerning this invention (claim 2) presumes one or more users' attribute information, A contents decision means of voice synthesis to determine the contents which speak based on said attribute information, An voice synthesis parameter decision means to determine the voice synthesis parameter for adjusting at least one side of the tune of wording and voice based on said attribute information, It is characterized by providing an voice synthesis means to utter facing uttering said determined contents and changing at least one side of the tune of wording and voice according to said determined voice synthesis parameter.

[0013] Preferably, said attribute information may include at least one information in a user's sex, age, and psychic distance. Preferably, said voice synthesis parameter may consist of at least one amount in the tenderness of the tune of voice, and the rate of voice synthesis whenever [ flank / of wording ].

[0014] In order to change one [ at least ] inclination of the tune of wording and voice into arbitration, you may make it hold preferably the internal rule which is used in case said voice synthesis parameter decision means determines said voice synthesis parameter from said attribute information and which can be changed.

[0015] The voice synthesis approach concerning this invention (claim 6) is characterized by uttering presuming one or more users' attribute information, determining the voice synthesis parameter for adjusting at least one side of the tune of wording and voice based on said attribute information, facing uttering the predetermined contents, and changing at least one side of the tune of wording and voice according to said determined voice synthesis parameter.

[0016] The voice synthesis approach concerning this invention (claim 7) presumes one or more users' attribute information. Based on said attribute information, determine the contents which speak, determine the voice synthesis parameter for adjusting at least one side of the tune of wording and voice based on said attribute information, and it faces uttering said determined contents. It is characterized by uttering changing at least one side of the tune of wording and voice according to said determined voice synthesis parameter.

[0017] Because according to this invention presume a user's age / sex / psychic distance and a machine changes the tune of wording or voice on the occasion of the dialogue by natural language with a user according to this presumed result Also when equipment faces only one user and coincidence is faced [ not to mention ] with two or more users Each user makes each user know whom the machine is speaking to according to the wording of equipment, or the difference of the tune of voice. Or felicity is sensed for this user with the tune of the wording which this user was made to have a sense of closeness over a machine with the tune of the wording towards a specific user, or voice, or was accordant to a specific user's psychic distance, or voice, and it can cheat.

[0018] Moreover, a user's age / sex / psychic distance are presumed, and it also becomes possible to also change the contents which a machine utters according to this presumed result. Therefore, each user can sense and have felicity in the tune of the wording which could know whom the machine would speak to according to the wording of equipment, or the difference of the tune of voice, and has had a sense of closeness to a machine by the ability addressing the contents for oneself at the tune of the wording and the voice only suitable for oneself, or was accordant to psychic distance with self, or voice, and can get a good feeling of use.

[0019] Again. In this invention, it makes it possible to make different characterization for every user interface perform easily by [ of the above-mentioned wording or the tune of voice ] changing and changing the inclination of the direction variously.

[0020] Therefore, for an equipment management person, it becomes possible by the above-mentioned wording and the tune of voice changing and changing the inclination of the direction variously to characterize every equipment systematically.

[0021] In addition, invention concerning each equipment [ more than ] is materialized also as invention concerning an approach, and invention concerning an approach is materialized also as invention concerning equipment. Moreover, the above-mentioned invention is materialized also as a medium which recorded the program for making a computer perform a corresponding procedure or a corresponding means and in which machine read is possible.

[0022]

[Embodiment of the Invention] Hereafter, the gestalt of implementation of invention is explained, referring to a drawing. Below, as the voice synthesis equipment which applied this invention, or an voice synthesis function, for example, equipment equipped with a user interface which performs a dialogue with human being who made voice the subject, or the active voice synthesis for human being, a multi-modal interface, etc., for example, information offer equipment, This operation gestalt is explained taking the case of the thing supposing what (or let the voice synthesis equipment which

applied this invention be information offer equipment, agent equipment, active advertising equipment, etc.) is mounted in agent equipment, active advertising equipment, etc.

[0023] This voice synthesis equipment or an voice synthesis function can be realized using software, and the various equipments or service which carried the voice synthesis function concerning this invention can be realized by mounting the program which realized this voice synthesis equipment or an voice synthesis function in a computer or various equipments.

[0024] First, the basic configuration of the voice synthesis equipment concerning 1 operation gestalt of this invention is explained. The basic configuration of the voice synthesis equipment applied to this operation gestalt at drawing 1 is shown. This voice synthesis equipment is equipped with the user description presumption section 1, the contents decision section 2 of voice synthesis, the voice synthesis parameter decision section 3, and the voice synthesis section 4.

[0025] If the user description presumption section 1 is required, it distinguishes "the person identification information (the identification information which shows a strange person is included)" showing the "person positional information" showing each user's location, each user's identifier, etc., while it detects a user separately and presumes "person attribute information", such as sex / age / psychic distance, for every user.

[0026] The contents decision section 2 of voice synthesis generates the "voice synthesis candidate information" and the "contents information of voice synthesis" that it means what is told to which user based on all each user's person identification information and person attribute information which were distinguished by the user description presumption section 1, person positional information, or its part.

[0027] The voice synthesis parameter decision section 3 computes the voice synthesis parameter information which controls the "wording" and/or "the tune of voice" according to person attribute information of the voice synthesis candidate presumed by the user description presumption section 1.

[0028] The voice synthesis section 4 if according to the voice synthesis candidate information determined in the contents decision section 2 of voice synthesis -- the face sense of an agent image, and a look (case the amount of [ about bodies of equipment equivalent to the doll which accompanies equipment, or this, such as a thing, the display screen, etc. ] moving part is electronically controllable -- face sense, such as a doll, -- a look) Or while controlling the confrontation directions, such as a body of equipment, and the display screen, etc., according to the voice synthesis parameter information determined in the voice synthesis parameter decision section 3, the voice synthesis output of the contents determined in the contents decision section 2 of voice synthesis is carried out.

[0029] Next, the fundamental processing in the voice synthesis equipment concerning this operation gestalt is explained. An example of the flow of the processing in the voice synthesis equipment applied to this operation gestalt at drawing 2 is shown. In the user description presumption processing of step S1, while presuming person attribute information, such as sex / age / psychic distance, for every user, if , the person identification information (the identification information which shows a strange person is included) showing the person positional information showing each user's location, each user's identifier, etc. is distinguished.

[0030] In the contents decision processing of voice synthesis of step S2, the voice synthesis candidate information and the contents information of voice synthesis that it means what is told to which user based on all each user's person identification information and person attribute information which were distinguished at step S1, person positional information, or its part are generated.

[0031] The voice synthesis parameter information which controls by voice synthesis parameter decision processing of step S3 the tune of the wording and/or the voice according to an voice synthesis candidate's person attribute information presumed at step S1 is computed.

[0032] In voice synthesis processing of step S4 if according to the voice synthesis candidate information determined at step S2 -- the face sense of an agent image, and a look (case the amount of [ about bodies of equipment equivalent to the doll which accompanies equipment, or this, such as a thing, the display screen etc. ] moving part is electronically controllable -- face sense, such as a doll, -- a look) Or while controlling the confrontation directions, such as a body of equipment, and the display screen, etc., according to the voice synthesis parameter information determined at step S3, the voice synthesis output of the contents determined at step S2 is carried out.

[0033] Below, it explains using the more detailed example of a configuration in order of the user description presumption section 1, the contents decision section 2 of voice synthesis, the voice synthesis parameter decision section 3, and the voice synthesis section 4. First, it explains in more detail about the user description presumption section 1.

[0034] The example of a configuration of the user description presumption section 1 is shown in drawing 3 . As shown in drawing 3 , this user description presumption section 1 is equipped with the person detection dictionary 11, the person detecting element 12, the person collating dictionary 13, the person collating section 14, the psychic distance presumption section 15, the sex collating dictionary 16, the sex presumption section 17, the age collating dictionary 18, and the age

presumption section 19.

[0035] The person detection dictionary 11 stores the template for person detection. The person detecting element 12 inputs the image of the space which faces the user of the front face of equipment, and it outputs a person's detected image part as a person field image while detecting a fully similar part with the above-mentioned template for person detection out of this image and outputting as person positional information.

[0036] The person collating dictionary 13 stores the template for person collating for every existing acquaintance object for collating an individual. The person collating section 14 collates with the above-mentioned template for person collating the person field image obtained by the person detecting element 12, chooses the highest and a fully similar template, and outputs the person identification information matched with this template. In addition, the person collating section 14 generates the person identification information which shows that he is a strange person, when a fully similar template is not able to be discovered, as a result of collating a person field image with the template for person collating.

[0037] The psychic distance presumption section 15 computes and outputs the psychic distance according to the this updated equipment use count while updating the equipment use count of the equipment concerned for every existing acquaintance object based on the above-mentioned person identification information. This psychic distance is a value which shows which was familiar with equipment in this case, and is the amount which carries out monotone reduction to an equipment use count, for example, the amount in inverse proportion to an equipment use count.

[0038] The sex collating dictionary 16 stores the template for sex collating. The sex presumption section 17 chooses the template which collates with the above-mentioned template for sex collating the person field image obtained by the person detecting element 12, and is similar to the highest, and outputs the sex matched with this template.

[0039] The age collating dictionary 18 stores the template for age collating. The age presumption section 19 chooses the template which collates with the above-mentioned template for age collating the person field image obtained by the person detecting element 12, and is similar to the highest, and outputs the age matched with this template.

[0040] In addition, like the person collating section 14, the sex presumption section 17 outputs the sex matched with this template, only when the highest and a fully similar template are able to be discovered, and when a fully similar template is not able to be discovered, you may make it output the information which shows that it is a sex indeterminate. The age presumption section 19 outputs the age matched with this template, only when the highest and a fully similar template are able to be discovered, and when a fully similar template is not able to be discovered, you may make it similarly output the information which shows that it is an age indeterminate.

[0041] In addition, they are the psychic distance which the psychic distance presumption section 15 outputs, the sex which the sex presumption section 17 outputs, and the age, i.e., person attribute information, which the age presumption section 19 outputs. Moreover, the template stored in the 16/age collating dictionary 18 of 13/sex collating dictionaries of person detection dictionaries of 11/person collating dictionaries is the information expressing the description of each category for which every [ which corresponds respectively ] category (according to the /man and woman classified by human being/individual / age) is statistically asked from the former image by which number acquisition was carried out enough.

[0042] Next, it explains in more detail about the contents decision section 2 of voice synthesis. The example of a configuration of the contents decision section 2 of voice synthesis is shown in drawing 4 . As shown in drawing 4 , this contents decision section 2 of voice synthesis consists of the control regulation database 21, the voice synthesis control section 22, a selection-rule database 23, a subject database 24, and the contents selection section 25 of voice synthesis.

[0043] The control regulation database (control regulation DB) 21 stores the control regulation for controlling the condition of voice synthesis. The voice synthesis control section 22 generates the contents selection command which controls the condition of voice synthesis according to the above-mentioned control regulation. A contents selection command expresses comparatively rough directions of the contents which should speak [ describe / "greet Mr. OO", / "describe main subject" ].

[0044] The selection-rule database (selection rule DB) 23 stores the selection rule for determining the contents which should speak. The subject database (subject DB) 24 stores the subject which should speak.

[0045] The contents selection section 25 of voice synthesis generates the voice synthesis candidate information which shows about which user it speaks as a target while reading the contents corresponding to the contents selection command which the voice synthesis control section 22 outputs from subject DB24 based on the selection rule in which it is stored by selection rule DB23.

[0046] Here, the control regulation stored in the control regulation DB21 is explained. If a control regulation is divided roughly, it will consist of two kinds, voice synthesis formation rule and the regulation for voice synthesis. The former is a regulation which manages advance of voice synthesis and the latter is a regulation which manages an voice synthesis candidate's decision.

[0047] First, voice synthesis formation rule are explained in detail. voice synthesis formation rule are regulations to which the condition of voice synthesis is made to change after the equipment concerned detects a user first until it finishes speaking predetermined subject, and lay down the voice synthesis sequence of a greeting or main subject.

[0048] An example of voice synthesis formation rule is shown in drawing 5 . In the voice synthesis formation rule illustrated to drawing 5 , it awaits, and from a condition, a user is detected, it is not rich, an initiation greeting ("hello" etc.) is uttered, and then the sequential voice synthesis of the predetermined main subject (some promotion, self-introduction, etc.) is carried out to this user (or a user's ensemble). voice synthesis in the main question is repeatedly performed until the main subject which should talk becomes that there is nothing in \*\*. If it finishes speaking about all main subject, a termination greeting ("thank you very much" etc.) will be uttered, and voice synthesis will be completed. If a user stops there being on the way, voice synthesis in the main question is stopped at the time, it will be uttered ("is it busy?"), an interruption greeting will await, and it will change in the condition.

[0049] Then, the regulation for voice synthesis is explained in detail. the regulation for an voice synthesis be a regulation which determine how the partner ( voice synthesis candidate ) who address in the midst of the voice synthesis which voice synthesis formation rule direct be change , be a clause unit or a sentence unit , and according to a candidate location and attribute information , it be use in order to change a look / face sense , such as an agent image , the tune of the wording/voice in the case of an voice synthesis , etc.

[0050] Two of round regulations are prepared for the regulation for voice synthesis a round regulation and the middle stage in which it is applied in the middle of the voice synthesis which continues after that at the time of the initiation applied at the time of voice synthesis initiation. For example, at the time of the initiation illustrated next, like a round regulation, it greets sequentially from the person located most in near, and the last is patrolled at random. When there is a child, the candidate for voice synthesis is made to patrol from a child preferentially.

[0051] An example of the initiation round regulation of the regulations for voice synthesis is shown in drawing 6 . At the time of the initiation illustrated to drawing 6 , under a round regulation, when there is a two parents-and-children companion's user, it greets as "hello and me" to a child first, and, subsequently it is introduced to parents, saying, "I am xx." In addition, the round is made into the sentence unit in this example. Although the number of greeting sentences is free, even if it finishes saying a greeting sentence, when there is a non-gone round round user further, application of a round regulation is ended at the time of initiation, and it goes into main subject.

[0052] Now, if it goes into main subject, a round regulation will be applied in the middle stage. An example of a round regulation is shown in drawing 7 in the middle stage among the regulations for voice synthesis. Under a round regulation, it addresses preferentially to the person who suits subject, and a person eager for subject the middle stage in which it illustrated to drawing 7 . Since there is a category (sex, age) of the person who was most suitable for subject respectively, this person's goodness of fit (the suiting number of attributes) is calculated by how many suit subject among a certain person's attribute information. Moreover, when a certain person tries positive approach to the equipment concerned or has turned the face to the equipment concerned, whenever [ eager ] (0/1) is calculated by judging that this person is an eager person. Next, a weighted sum a goodness of fit and whenever eager is made into a round priority, and a candidate is patrolled according to this round priority. The time amount for one person is about dozens of seconds from several seconds, and is proportional to a round priority.

[0053] Next, the selection rule stored in selection rule DB23 are explained. Selection rule are regulations for choosing the contents of voice synthesis according to a contents selection command. An example of selection rule is shown in drawing 8 .

[0054] In the selection rule illustrated to drawing 8 , although contents with an initiation greeting, a termination greeting, and an interruption greeting common to an adult/child are chosen, about main subject, contents which are different by the adult and the child are chosen. In addition, sex and distinction by psychic distance are also considered as selection rule in the main question.

[0055] Next, the information stored in subject DB24 is explained. The contents of subject DB24 are the texts described in standard wording. The contents of the subject DB24 at the time of applying the voice synthesis equipment concerning this operation gestalt as active advertising equipment are illustrated below.

[0056] [The example of subject]

object = -- existing acquaintance object: -- classification = initiation greeting: -- sentence = -- "-- hello -- Mr. OO -- "

object = -- strange -- person:classification = initiation greeting: -- sentence = -- "how do you do. I am xx. "

object = -- all category: -- classification = termination greeting: -- sentence = "thank you very much"

object = -- all category: -- classification = interruption greeting: -- sentence = -- "-- is it busy? Let's meet you again. "

object = adult: -- classification = main subject: -- sentence = "now, recently --" (guidance of traveling abroad)

object = adult: -- classification = main subject: -- sentence = "it is -- most now" (guidance of a new work movie)

object = child: -- classification = main subject: -- sentence = "next time --" (guidance of a new work game)  
 object = adult: -- classification = main subject: -- sentence = "it is -- most now" (guidance of a new work movie)  
 object = -- all category: -- classification = main subject: -- sentence = "new attraction --" (guidance of a theme park)  
 object = adult and female: -- classification = main subject: -- sentence = "this fresh color of autumn --" (guidance of cosmetics)

Each subject consists of the candidate list which specifies a candidate's category, identification information which specifies the classification of subjects, such as a greeting and main subject, and an actual sentence. The contents selection section 25 of voice synthesis searches the subject which contains in a candidate list the candidate of the subject defined according to application of selection rule, and it works so that the subject which has not completed voice synthesis may be selected one by one. The sentence indicated by the selected subject is the contents information of voice synthesis which the contents decision section 2 of voice synthesis outputs.

[0057] however, the sentence which is used as an object = existing acquaintance object in the case of a strange person, for example, an object, -- = existing acquaintance object:classification = initiation greeting:sentence = -- "-- hello -- Mr. OO - -"

\*\*\*\*\* is made improper.

[0058] In addition, when supplementary data (catalog photograph etc.) are linked to a proper place and a link part is uttered in the sentence of subject DB24, the corresponding supplementary data may be made to be shown to coincidence. Moreover, the information which specifies body actuation of the agent at that time is also added to each sentence, for example, these data point at the time of salutation actuation and supplementary data presentation at the time of a greeting, and an agent may enable it to perform actuation etc.

[0059] Next, it explains in more detail about the voice synthesis parameter decision section 3. The example of a configuration of the voice synthesis parameter decision section 3 is shown in drawing 9. As shown in drawing 9, this voice synthesis parameter decision section 3 consists of the standard property database 31, the canonical parameter setting section 32, an individual property database 33, and the individual parameter setup section 34.

[0060] The standard property database (standard property DB) 31 stores the standard adjustment set for determining a standard voice synthesis parameter according to a candidate's category. The canonical parameter setting section 32 determines standard voice synthesis parameter information based on the person attribute information which the user description presumption section 1 outputs, and the above-mentioned standard adjustment set.

[0061] The individual property database (individual property DB) 33 stores the individual property for expressing the character of equipment each. the individual parameter setup section 34 receives the canonical parameter information which the canonical parameter setting section 32 outputs, and is individual to this -- the voice synthesis parameter information that the deviation by the individual property stored in DB33 was added is generated.

[0062] The standard property DB31 has one fundamental parameter set, and the standard adjustment set for every category of the candidate who specified the amount of adjustments to this is stored. Specifically, the standard adjustment set for children, the standard adjustment set for women, etc. are stored. The canonical parameter setting section 32 reads the standard adjustment set which corresponds from a candidate's given category, adjusts a fundamental parameter set based on this, and generates the canonical parameter according to a category.

[0063] The example of a fundamental parameter set and a standard adjustment set is shown below.

[fundamental parameter set] whenever [ flank / of wording ] -- =50 (certified value) whenever [ tenderness / of the tune of voice ] -- =50 (certified value) Speed =50 about which it speaks (certified value) [Standard adjustment set for children] Whenever [ flank / of wording ] -- +50 (it adjusts to relaxed and informal sensibility extremely) whenever [ tenderness / of the tune of voice ] +50 (it adjusts to sensibility very gentle to) Speed about which it speaks -30 (it adjusts to the sensibility carried out slowly) [Standard adjustment set for adults] Whenever [ flank / of wording ] \*\*0 (certified value maintenance) whenever [ tenderness / of the tune of voice ] \*\*0 (certified value maintenance) speed about which speak \*\*0 (certified value maintenance) [the standard adjustment set for old men] Whenever [ flank / of wording ] \*\*0 (certified value maintenance) Whenever [ tenderness / of the tune of voice ] +20 (it adjusts to sensibility a little gentle to) Speed about which it speaks -30 (it adjusts to the sensibility carried out slowly) [the near standard adjustment set for persons of psychic distance] whenever [ flank / of wording ] +50 (it adjusts to relaxed and informal sensibility extremely) Whenever [ tenderness / of the tune of voice ] \*\*0 (certified value maintenance) Speed about which it speaks \*\*0 (certified value maintenance) [Far standard adjustment set for persons of psychic distance] Whenever [ flank / of wording ] () -50 It adjusts to the sensibility were extremely changed. Whenever [ tenderness / of the tune of voice ] \*\*0 (certified value maintenance) Speed about which it speaks \*\*0 (certified value maintenance) These are the standard wording decided for every category of the candidate also common to which equipment, and a regulation which sets up the tune of voice. It changes into the wording which was able to give the individuality for every equipment for this by the



individual parameter setup section 34 of the next step, and the parameter which can generate the tune of voice. The individual property for attaching this individuality is stored in the individual property DB33.

[0064] The example of an individual property is shown below.

[Individual property common to a category] (as opposed to whom) Whenever [ flank / of wording ] +10 it is \*\*\*\* -- inclination to talk by beam sensibility (as opposed to whom) Whenever [ tenderness / of the tune of voice ] +10 \*\* -- inclination to talk gently Speed, about which it speaks -20 (inclination about which it speaks slowly to anyone) [category -- individual property] of a proper if (a candidate is a child) then Tenderness of the tune of voice +10 (inclination especially gentle to a child)

An individual property common to a category is the individuality of the equipment independent of a candidate's category, and means talking with the same inclination to anyone. The individual property of a proper demonstrates effect to a candidate's specific category like equipment gentle to a child in a category. Thus, it becomes possible to direct the individuality of equipment by the common framework by establishing an individual property.

[0065] Next, it explains in more detail about the voice synthesis section 4. The example of a configuration of the voice synthesis section 4 is shown in drawing 10 . As shown in drawing 10 , this voice synthesis section 4 consists of the sentence transformation-rule database 41, an voice synthesis sentence transducer 42, the acoustic-sense information generation section 43, and the vision information generation section 44.

[0066] The sentence transformation-rule database (sentence transformation rule DB) 41 stores the regulation for changing into the sentence of predetermined wording the contents information of voice synthesis (it having been described by standard wording reading out sentence) which the contents decision section 2 of voice synthesis outputs.

[0067] It is changed into the sentence of predetermined wording, the voice synthesis sentence transducer 42 referring to the above-mentioned sentence transformation rule DB41 according to the voice synthesis parameter with which the contents decision section 2 of voice synthesis outputs the above-mentioned contents information of voice synthesis.

[0068] The acoustic-sense information generation section 43 carries out [ voice / actual ] the conversion output of the voice synthesis sentence information which the voice synthesis sentence transducer 42 outputs according to the above-mentioned voice synthesis parameter. the vision information generation section 44 generate the face sense / look according to voice synthesis candidate information , and the agent image which be able to give the actuation according to the additional information of the contents information of an voice synthesis ( or -- electronic -- a controllable doll etc. -- control ) , and if required , it convert into a video signal in coincidence the supplementary data linked to the contents information of an voice synthesis .

[0069] The acoustic-sense information generation section 43 delivers the synchronization information corresponding to this voice to the vision information generation section 44 while carrying out [ voice ] a generation output. The vision information generation section 44 which received synchronization information and was passed opens and closes an agent's opening according to this synchronization information.

[0070] In addition, the relation between voice synthesis candidate information, voice synthesis parameter information, the contents information of voice synthesis, and the voice synthesis sentence information changed based on them is as follows, for example at this time.

[The example of voice synthesis candidate information]

Name = Suzuki scent sex = male age =8 years-old psychic distance = 60 locations =xx [the example of voice synthesis parameter information]

Tenderness of the tune of degree =of flank 100 voice of wording = example] of the speed =40[voice synthesis sentence information (before conversion) about which it speaks 100

"hello -- Mr. OO. I introduce -- in which a new attraction was substantial. "

[The example of voice synthesis sentence information (after conversion)]

"hello -- the scent. Does a new attraction introduce full --? "

In order to perform this, the following transformation rules are described by the sentence transformation rule DB41. to transformation rule, it illustrates below -- as -- a name substitution regulation and the sentence end -- there are transformation rule and a paraphrasing regulation. An example of a name substitution regulation is shown in drawing 11 . drawing 12 -- the sentence end -- an example of transformation rule is shown. It is put in another way as drawing 13 , and an example of a regulation is shown.

[0071] The above-mentioned name substitution regulation is a regulation for substituting a name suitable for this part, when a candidate's etc. identifier must be put into the contents information of voice synthesis. moreover, the sentence end -- transformation rule are regulations for deflecting the sentence end appropriately according to whenever [ flank ]. A paraphrasing regulation is a regulation for transposing a word difficult especially for a child to a plain word.

[0072] the sentence end -- it must be put in another way as transformation rule, and a regulation must inspect whenever

[ flank / of each word used in the text (the contents information of voice synthesis) ], and difficulty on the occasion of the application. An allowed value list is used for this inspection. An allowed value list is information which specifies the applicability of whenever [ flank / of each word ], or difficulty. The word which has deviated from this range must be transposed to another word from which it does not deviate. A synonym list is used for this replacement. A synonym list is the information which summarized the synonym of each word. The word of the text which deviated from the allowed value is searched from this synonym list, and another word from which it does not deviate is replaced. When two or more synonyms which can be replaced exist, one is decided and replaced at random. It avoids that the same word appears by doing in this way repeatedly in a sentence.

[0073] The example of an allowed value list and a synonym list is shown below. In addition, an allowed value list and a synonym list are stored in the sentence transformation rule DB41 with sentence transformation rule.

[The example of an allowed value list]

∴, such as "it carries out", whether "whether it to carry out", "it carries out", and "not carrying out" flank whenever > -- :80 >= flank whenever >=, such as 80 "it carries out" and "not carrying out", -- 20 "I will carry out", "I doing", and "I not doing" -- "-- " (prefix) -- etc. -- whenever [ :20> flank ] -- "-- :difficulty >, such as fullness", -- :50 >= difficulty, such as that it is 50 "full" and "it being full", -- "-- how do you do :difficulty >, such as ", :50, such as 50 "hello and me", >= difficulty [the example of a synonym list]

The voice generation section 43, such as "it carries out", whether "whether it to carry out", "it carries out", "it carrying out", "I carrying out", "I doing", etc. "it not carrying out" "it not carrying out", and "not doing", changes into voice the voice synthesis sentence information generated as mentioned above with the tune and speed of voice which an voice synthesis parameter specifies. each sentence end -- the information on taking up and down of the ending is attached to the word. For example, the ending of "whether to carry out" goes up and the ending of "carrying out" falls. in addition, when the tune of voice is gentle, the dynamic range of speech power is narrowed and is attached -- not being audible -- making -- further -- the sentence end -- the processing which extends the ending of a word is added and tenderness is expressed.

[0074] In addition, voice synthesis equipment or the voice synthesis approach concerning this invention are not limited to the above operation gestalt. For example, the contents of voice synthesis are not changed according to a user's category, but you may make it utter the predetermined contents at the tune of the wording and the voice according to a user.

[0075] Moreover, it is good also as adjustment being variously possible by the actuation from the outside as the equipment exterior to exchange of the various information stored in the person detection dictionary 11, the person collating dictionary 13, the sex collating dictionary 16, the age collating dictionary 18, the control regulation DB21, selection rule DB23, subject DB24, the standard property DB31, the individual property DB33, and the sentence transformation rule DB41 being possible.

[0076] Moreover, the user description presumption section 1 is possible also for independent or using for coincidence and presuming person identification information, person positional information, and person attribute information in other information, such as not only an image but voice.

[0077] Moreover, it is good also as possible in the output which changed wording not in voice but in the alphabetic character. By the way, it is also possible to apply the information which recorded on the record medium 51 and this recorded the information (for example, program) which realizes the voice synthesis equipment concerning this invention and the voice synthesis approach as shown in drawing 14 to equipment 53 via a communication line in to apply to equipment 52 \*\*\*\* via this record medium 51. This invention is not limited to the gestalt of operation mentioned above, in the technical range, can deform variously and can be carried out.

[0078]

[Effect of the Invention] When equipment faces two or more users at coincidence according to this invention, each user can know whom the machine is speaking to according to the wording of equipment, or the difference of the tune of voice. By the ability addressing the contents for oneself at the tune of the wording and the voice only suitable for oneself, it can sense and have felicity in the tune of the wording which has had a sense of closeness to a machine, or was accordant to psychic distance with self, or voice, and a good feeling of use can be obtained.

[0079] Moreover, for an equipment management person, it becomes possible by the above-mentioned wording and the tune of voice changing and changing the inclination of the direction variously to characterize every equipment systematically.

---

[Translation done.]



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**